Distributed Diagnosis of Actuator and Sensor Faults in HVAC Systems ⋆

Panayiotis M. Papadopoulos * **Vasso Reppa** *
Marios M. Polycarpou * **Christos G. Panayiotou** *

* *KIOS Research and Innovation Center of Excellence*
Department of Electrical and Computer Engineering
University of Cyprus, Nicosia, 1678, Cyprus
(e-mail: papadopoulos.panagiotis@ucy.ac.cy, reppavasso@gmail.com,
mpolycar@ucy.ac.cy, christosp@ucy.ac.cy).

Abstract: This paper presents a model-based methodology for diagnosing actuator and sensor faults affecting the temperature dynamics of a multi-zone heating, ventilating and air-conditioning (HVAC) system. By considering the temperature dynamics of the HVAC system as a network of interconnected subsystems, a distributed fault diagnosis architecture is proposed. For every subsystem, we design a monitoring agent that combines local and transmitted information from its neighboring agents in order to provide a decision on the type, number and location of the faults. The diagnosis process of each agent is realized in three steps. Firstly, the agent performs fault detection using a distributed nonlinear estimator. After the detection, the local fault identification is activated to infer the type of the fault using two distributed adaptive estimation schemes and a combinatorial decision logic. In order to distinguish between multiple local faults and propagated sensor faults, a distributed fault isolation is applied using the decisions of the neighboring agents. Simulation results of a 5-zone HVAC system are used to illustrate the effectiveness of the proposed methodology.

Keywords: actuator faults, sensor faults, building automation, distributed fault diagnosis

1. INTRODUCTION

Due to the incessant operation of HVAC systems, which is necessary for satisfying the day-and-night demands of occupants for home comfort, the HVAC equipment may gradually or suddenly fail. The abnormal behavior of the HVAC components may cost the occupants comfort and money, while making the overall system unreliable. Large and abrupt faults (also called hard faults) can mostly be noticed by the occupants or the technical personnel. On the contrary, soft faults do not usually manifest themselves in an easily observable way by humans. This fact does not imply that soft faults are less hazardous than hard faults since this class of faults can lead the system to a long-term penetration and over-consumption. The application of automated fault diagnosis (FD) mechanisms has been proposed as a reliable solution to avoid a number of major consequences for buildings such as system shut-down [Liang and Du (2007); Wang et al. (2010)].

FD techniques for HVAC systems can be categorized according to the type of possible faults affecting the system. Several researchers have proposed FD schemes that tackle the problem of process faults [Bonvini et al. (2014)], or actuator faults [Weimer et al. (2012); Carbot-Rojas et al. (2015); Darure et al. (2016)] or sensor faults [Wang et al. (2010); Reppa et al. (2015)]. In practice however it is not known a priori what type of fault will occur. There are very few FD methods that address the problem of heterogeneous faults for both actuator and sensor faults [Lee et al. (2004)].

In the last decade, there has been a significant research activity in the design of non-centralized FD architectures [Blanke et al. (2016); Reppa et al. (2016); Boem et al. (2017); Keliris et al. (2015)]. Decentralized and distributed methods are considered more suitable for large-scale applications than the centralized methods, especially with respect to scalability (e.g., installation of a new fan coil unit to an existing system) and security (e.g., a malicious attack of the central Building Management System). In previous work, the authors have developed a distributed detection and isolation approach for sensor faults affecting the electromechanical part and the non-interacting zones of a HVAC system [Reppa et al. (2015)]. Recently, the authors have designed a distributed adaptive estimation scheme for diagnosing sensor faults and accommodating their effects in a network of interacting HVAC building zones [Papadopoulos et al. (2015a,b)].

The goal and the main contribution of this work is the design of a model-based distributed FD architecture for isolating bias sensor and actuator faults in a multi-zone HVAC system that is considered as a network of interconnected subsystems as presented in Section 2. The proposed FD scheme uses several distributed monitoring agents, where every agent combines local and neighboring information to diagnose faults in its monitored subsystem. The monitoring agent performs a sequence of diagnostic processes, including: (i) distributed fault detection for

⋆ This work was supported by the European Research Council under the ERC Advanced Grant ERC-2011-AdG-291508

capturing the occurrence of faults in the monitored subsystem and its neighborhood (Section 3.1), (ii) local fault identification for specifying the type of local faults, i.e., actuator or sensor fault (Section 3.2), and (iii) distributed fault isolation for isolating multiple local faults and/or propagated sensor faults (Section 3.3). The proposed FD scheme is applied to a 5-zone HVAC system, presented in Section 4.

2. PROBLEM FORMULATION

Let us consider a multi-zone HVAC system as shown in Fig. 1, which consists of the electromechanical part (green box) and N building zones interconnected through doors and walls [Witrant et al. (2009)]. The multi-zone HVAC system is regarded as a network of $N+1$ interconnected subsystems Σ^s, $\Sigma^{(1)}$,...,$\Sigma^{(N)}$, where Σ^s represents the temperature dynamics of the storage tank and $\Sigma^{(i)}$, $i \in \{1,\dots,N\}$ represents the temperature dynamics of a building zone [Papadopoulos et al. (2015b)].

The subsystem Σ^s can be expressed as

$$\begin{aligned}\Sigma^s:\ \dot{x}^s(t) =& A^s x^s(t) + g^s(x^s(t), d^s(t))u^s(t) \\ &+ h^s(x^s(t), x(t), u(t)) + \eta^s(d^s(t)) + r^s(t),\end{aligned} \tag{1}$$

where $x^s \in \mathbb{R}$ represents the water temperature of the storage tank (system state) and $u^s \in \mathbb{R}$ denotes the normalized energy in the heat pump (control input). The vector $d^s \triangleq [d_1^s, d_2^s]^\top$ represents an uncontrollable but known exogenous input vector, where d_1^s is plenum (duct) temperature and d_2^s is the source heat temperature to the heat pump. The variable $r^s \in \mathbb{R}$ models unknown disturbances affecting the water temperature dynamics due to e.g. defective thermal insulation of the storage tank. The vector $x \triangleq [x^{(1)}, \dots, x^{(N)}]$ is the interconnection vector, where $x^{(i)}$ is the air temperature of the interconnected building zone i (i.e. state of subsystem $\Sigma^{(i)}$), and $u \triangleq [u^{(1)}, \dots, u^{(N)}]$, where $u^{(i)}$ is the mass flow rate of water flowing into the fan coil of the ith building zone (i.e. control input of subsystem $\Sigma^{(i)}$). The terms $g^s \in \mathbb{R}$ and $h^s \in \mathbb{R}$ are defined as,

$$g^s(x^s, d^s) = \frac{U_{st,max}}{C_{st}}\left(1 + p\left(1 - \frac{x^s(t) - d_2^s(t)}{\Delta T_{max}}\right)\right), \tag{2}$$

$$h^s(x^s, x, u) = \frac{a_{sz}}{C_{st}} \sum_{i\in\{1,\dots,N\}} U_{i,max}(x^s - x^{(i)})u_c^{(i)}, \tag{3}$$

while the constant A^s is defined as $A^s = -\frac{a_{st}}{C_{st}}$ and $\eta^s(d^s) = \frac{a_{st}}{C_{st}} d_1^s$. Note that under healthy and faulty conditions $x^s(t) \le \Delta T_{max} + d_s^s(t)$. The constant parameters of Σ^s

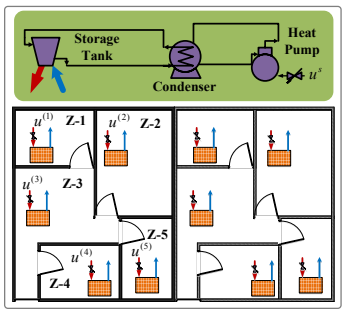

Fig. 1. Schematic representation of a multi-zone HVAC system. The orange rectangular boxes represent the fan-coil units.

are described in [Papadopoulos et al. (2015b)]. The state of Σ^s(water temperature in storage tank) is measured by the sensor $\mathcal{S}^s$, characterized by

$$\mathcal{S}^s : y^s(t) = x^s(t) + n^s(t) \tag{4}$$

where $y^s \in \mathbb{R}$ is the sensor output and $n^s \in \mathbb{R}$ is the measurement noise.

Let us define the set of indexes of subsystems $\Sigma^{(j)}$ that are interconnected with $\Sigma^{(i)}$ as $\mathcal{K}_i = \{j \in \{1,\dots,N\}\backslash\{i\} : a_{z_{ij}} \neq 0 \text{ or } A_{d_{ij}} \neq 0\}$, where $a_{z_{ij}}$ is the coefficient of the inter-zone thermal flow due to the wall between i-th and j-th zone, and $A_{d_{ij}}$ is the area of the door between i-th and j-th zone. Each subsystem $\Sigma^{(i)}$, $i \in \{1,\dots,N\}$ is interconnected with Σ^s and card$(\mathcal{K}_i)$ [1] subsystems $\Sigma^{(j)}$, $j \in \mathcal{K}_i$, $j \neq i$, described by

$$\begin{aligned}\Sigma^{(i)}:\ \dot{x}^{(i)}(t) =& A^{(i)}x^{(i)}(t) + g^{(i)}(x^s(t), x^{(i)}(t))u^{(i)}(t) \\ &+ h^{(i)}(x^{(i)}(t), x_{\mathcal{K}_i}(t)) + \eta^{(i)}(d^{(i)}(t)) + r^{(i)}(t),\end{aligned} \tag{5}$$

where $x_{\mathcal{K}_i}(t) = \left[x^{(j)}(t) : j \in \mathcal{K}_i\right]^\top$, $x_{\mathcal{K}_i}$ denotes a column vector of length card$(\mathcal{K}_i)$, where each element corresponds to the state $x^{(j)}$ of the neighboring subsystem $\Sigma^{(j)}$, $j \in \mathcal{K}_i$. The variable $r^{(i)}$ models the unknown system disturbances of the subsystem $\Sigma^{(i)}$, due to e.g. appliances, occupants, lights and $A^{(i)} = \frac{hA_{w_i} - a_{z_i}}{C_{z_i}} - \frac{1}{C_{z_i}} \sum_{j\in\mathcal{K}_i} a_{z_{ij}}$. The terms $g^{(i)} \in \mathbb{R}$ and $h^{(i)} \in \mathbb{R}$ are described by

$$g^{(i)}(x^s, x^{(i)}) = \sigma^{(i)}(x^s - x^{(i)}), \tag{6}$$

$$\begin{aligned}h^{(i)}(x^{(i)}, x_{\mathcal{K}_i}) =& \frac{1}{C_{z_i}} \sum_{j\in\mathcal{K}_i} a_{z_{ij}} x^{(j)} + p^{(i)}\Big(\sum_{j\in\mathcal{K}_i} sgn(x^{(j)} - x^{(i)}) \\ &\times A_{d_{ij}} \max(x^{(i)}, x^{(j)}) \sqrt{\left|x^{(j)} - x^{(i)}\right|}\Big)\end{aligned} \tag{7}$$

with $\sigma^{(i)} = \frac{U_{i,max}a_{sz}}{C_{z_i}}$, $p^{(i)} = \frac{\rho_{air}C_p\sqrt{2(C_p - C_v)}}{C_{z_i}}$, and $\eta^{(i)}(d^{(i)}) = \frac{a_{z_i}}{C_{z_i}} d_1^{(i)} - \frac{hA_{w_i}}{C_{z_i}} d_2^{(i)}$, where $d_1^{(i)}$ is the temperature of the surface node of the mass wall in the i-th zone and $d_2^{(i)}$ is the ambient temperature, respectively.

The objective of this work is to design a methodology for diagnosing actuator and sensor faults that may occur in one or more building zones, assuming that there are no actuator and sensor faults in the electromechanical part of HVAC. The output of the sensor $\mathcal{S}^{(i)}$ used to measure the state (air temperature in zone i) of subsystem $\Sigma^{(i)}$ is expressed as

$$\mathcal{S}^{(i)} : y^{(i)}(t) = x^{(i)}(t) + n^{(i)}(t) + f_s^{(i)}, \tag{8}$$

where $y^{(i)} \in \mathbb{R}$ is the sensor output, $n^s \in \mathbb{R}$ is the measurement noise and $f_s^{(i)} \in \mathbb{R}$ denotes the permanent bias sensor fault (modeled as in Reppa et al. (2016)). The input of $\Sigma^{(i)}$ is affected by actuator faults modeled as

$$u^{(i)}(t) = u_c^{(i)}(t) + f_a^{(i)}, \tag{9}$$

where $f_a^{(i)}$ is the actuator bias fault that may affect the valve regulating the flow of water in fan-coil unit of the i-th zone. The signals u^s in (1) and $u_c^{(i)}$ in (9) are generated using a distributed feedback linearization controller based on some (differentiable) reference signals y_{ref}^s and $y_{ref}^{(i)}$ for the states x^s and $x^{(i)}$, $i \in \{1,\dots,N\}$.

[1] card$(\cdot)$ denotes the cardinality of a set

3. DISTRIBUTED FAULT DIAGNOSIS

For the design of the fault diagnosis method for the multi-zone HVAC system described in the previous section we follow a distributed approach. Figure 2 illustrates the distributed diagnosis architecture for a simple example of two interconnected subsystems $\Sigma^{(1)}$ and $\Sigma^{(2)}$. For each subsystem, we design a monitoring agent $\mathcal{M}^{(i)}$, $i \in \{1,2\}$. The agent $\mathcal{M}^{(i)}$ exchanges information with its neighbor, where the exchange of information is coherent with the form of the physical interconnections. The diagnosis process is executed in three steps: distributed fault detection; local fault identification; and distributed fault isolation.

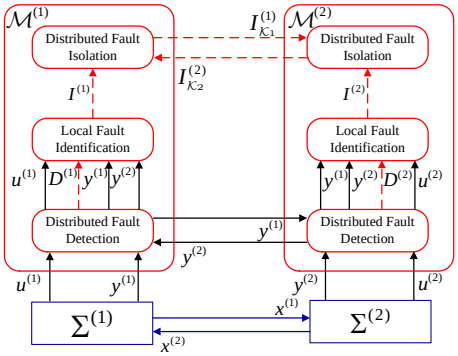

Fig. 2. Distributed fault diagnosis for two interconnected building zones

3.1 Distributed Fault Detection

By using the local input and sensor output information, $u^{(i)}$ and $y^{(i)}$, as well as sensor information transmitted by the neighboring agents $y^{(j)}$, $j \in \mathcal{K}_i$, the agent $\mathcal{M}^{(i)}$ detects the occurrence of faults that may have affected the local actuator or sensor, or may have been propagated due to the exchange of the sensor information. The fault detection decision logic of $\mathcal{M}^{(i)}$ is based on an analytical redundancy relation (ARR) $\mathcal{E}^{(i)}$, defined as

$$\mathcal{E}^{(i)}: \quad |\varepsilon_y^{(i)}(t)| \leq \bar{\varepsilon}_y^{(i)}(t), \tag{10}$$

where $\varepsilon_y^{(i)}(t)$ is the residual and $\bar{\varepsilon}_y^{(i)}(t)$ is the adaptive threshold, defined next. Under healthy conditions, i.e. when $f_a^{(i)} = 0$, $f_s^{(j)}=0$ for all $j \in \{\mathcal{K}_i \cup \{i\}\}$, $\mathcal{E}^{(i)}$ is guaranteed to be satisfied by designing the threshold $\bar{\varepsilon}_y^{(i)}(t)$ to bound the magnitude of the residual. Therefore, if there is a time instant that $\mathcal{E}^{(i)}$ is not satisfied, the agent $\mathcal{M}^{(i)}$ infers the presence of faults in its monitored subsystem $\Sigma^{(i)}$ and/or its neighbors. The output of the agent $\mathcal{M}^{(i)}$ is the boolean decision signal $D^{(i)}$, defined as

$$D^{(i)}(t) = \begin{cases} 0, & t < t_D^{(i)} \\ 1, & t \geq t_D^{(i)} \end{cases}, \tag{11}$$

where $t_D^{(i)} \triangleq inf\{t \geq 0 : |\varepsilon_y^{(i)}(t)| > \bar{\varepsilon}_y^{(i)}(t)\}$. When $D^{(i)}(t) = 1$, the agent $\mathcal{M}^{(i)}$ detects the occurrence of faults.

Residual Generation: The fault detection process of the agent $\mathcal{M}^{(i)}$ is executed by monitoring the residual

$$\varepsilon_y^{(i)}(t) = y^{(i)}(t) - \hat{x}^{(i)}(t) \tag{12}$$

where $\hat{x}^{(i)}$ is the estimation of $x^{(i)}$ generated by the following distributed nonlinear estimator

$$\begin{aligned}\dot{\hat{x}}^{(i)}(t) = A^{(i)}\hat{x}^{(i)}(t) + g^{(i)}(y^s(t), y^{(i)}(t))u_c^{(i)}(t) + \eta^{(i)}(d^{(i)}(t)) \\ + h^{(i)}(y^{(i)}(t), y_{\mathcal{K}_i}(t)) + L^{(i)}\left(y^{(i)}(t) - \hat{x}^{(i)}(t)\right),\end{aligned} \tag{13}$$

where $y_{\mathcal{K}_i}(t) = \left[y^{(j)}(t) : j \in \mathcal{K}_i\right]^\top$, $\hat{x}^{(i)}(0) = 0$ and $L^{(i)}$ is the observer gain selected such that $A_L^{(i)} = A^{(i)} - L^{(i)}$ is stable and y^s and $y^{(i)}$ are defined in (4) and (8).

Computation of Adaptive Threshold: The adaptive threshold $\bar{\varepsilon}_y^{(i)}$ is designed to bound the corresponding residual as shown in (10) when $f_a^{(i)}=0$, $f_s^{(j)}=0$ for all $j \in \{\mathcal{K}_i \cup \{i\}\}$. In this case, the residual is described by

$$\varepsilon_y^{(i)}(t) = \varepsilon_x^{(i)}(t) + n^{(i)}(t), \tag{14}$$

where $\varepsilon_x^{(i)} = x^{(i)} - \hat{x}^{(i)}$ is the state estimation error that satisfies,

$$\begin{aligned}\dot{\varepsilon}_x^{(i)} = A_L^{(i)}\varepsilon_x^{(i)} + \tilde{g}^{(i)}(x^s, x^{(i)}, y^s, y^{(i)})u_c^{(i)} \\ + \tilde{h}^{(i)}(x^{(i)}, x_{\mathcal{K}_i}, y^{(i)}, y_{\mathcal{K}_i}) + r^{(i)} - L^{(i)}n^{(i)}\end{aligned} \tag{15}$$

$$\tilde{g}^{(i)}(x^s, x^{(i)}, y^s, y^{(i)}) = g^{(i)}(x^s, x^{(i)}) - g^{(i)}(y^s, y^{(i)}) \tag{16}$$

$$\tilde{h}^{(i)}(x^{(i)}, x_{\mathcal{K}_i}, y^{(i)}, y_{\mathcal{K}_i}) = h^{(i)}(x^{(i)}, x_{\mathcal{K}_i}) - h^{(i)}(y^{(i)}, y_{\mathcal{K}_i}). \tag{17}$$

The adaptive threshold is computed by introducing the solution of (15) in (14) and bounding each term, taking into account the following assumptions:

Assumption 1: The measurement noise n^s, $n^{(i)}$ and the system disturbance $r^{(i)}$ are uniformly bounded; i.e. $|n^s(t)| \leq \bar{n}^s$, $|n^{(i)}(t)| \leq \bar{n}^{(i)}$, and $|r^{(i)}(t)| \leq \bar{r}^{(i)}$.

Assumption 2: The states x^s, $x^{(i)}$ and control inputs u^s, $u_c^{(i)}$, for all $i \in \{1, \ldots, N\}$ remain bounded under both healthy and faulty conditions; i.e., $x^s \in \mathcal{X}^s$, $x^{(i)} \in \mathcal{X}^{(i)}$ and $u^s \in \mathcal{U}^s$, $u_c^{(i)} \in \mathcal{U}^{(i)}$, where $\mathcal{X}^s$, $\mathcal{X}^{(i)}$, $\mathcal{U}^s$ and $\mathcal{U}^{(i)}$ are compact closed sets, respectively.

Based on (6) and (7), the bounds on the functions $\tilde{g}^{(i)}$, $\tilde{h}^{(i)}$ (see (16), (17)) are computed by setting $f_s^{(i)}=0$ in (8); i.e.,

$$|\tilde{g}^{(i)}(x^s, x^{(i)}, y^s, y^{(i)})| \leq \sigma^{(i)}\left(\bar{n}^{(i)} + \bar{n}^s\right) = \bar{g}^{(i)}(\bar{n}^{(i)}, \bar{n}^s), \tag{18}$$

$$\begin{aligned}|\tilde{h}^{(i)}(x^{(i)}, x_{\mathcal{K}_i}, y^{(i)}, y_{\mathcal{K}_i})| \leq p^{(i)} \sum_{j \in \mathcal{K}_i} A_{d_{ij}} \bar{\mu}^{(i)}(y^{(i)}, y^{(j)}) \\ + \sum_{j \in \mathcal{K}_i} \frac{a_{z_{ij}}}{C_{z_i}} \bar{n}^{(j)} = \bar{h}^{(i)}(y^{(i)}, y_{\mathcal{K}_i}),\end{aligned} \tag{19}$$

where $\bar{n}_{\mathcal{K}_i} = \left[\bar{n}^{(j)} : j \in \mathcal{K}_i\right]^\top$. The function $\bar{\mu}^{(i)}(y^{(i)}, y^{(j)})$ is computed such that $|\mu^{(i)}(x^{(i)}, x^{(j)}) - \mu^{(i)}(y^{(i)}, y^{(j)})| \leq \bar{\mu}^{(i)}(y^{(i)}, y^{(j)})$, where

$$\mu^{(i)}(x^{(i)}, x^{(j)}) = \text{sgn}(x^{(j)} - x^{(i)})\max(x^{(i)}, x^{(j)})\sqrt{|x^{(j)} - x^{(i)}|}. \tag{20}$$

The detailed computation of $\bar{\mu}^{(i)}(y^{(i)}, y^{(j)})$ is given in Appendix A.

The adaptive threshold $\bar{\varepsilon}_y^{(i)}(t)$ is defined as

$$\begin{aligned}\bar{\varepsilon}_y^{(i)} = \rho^{(i)}e^{-\lambda^{(i)}t}\bar{x}^{(i)} + \bar{n}^{(i)} + \int_0^t \rho^{(i)}e^{-\lambda^{(i)}(t-\tau)}\Big(\left|L^{(i)}\right|\bar{n}^{(i)} + \bar{r}^{(i)} \\ + \bar{g}^{(i)}(\bar{n}^{(i)}, \bar{n}^s)\left|u_c^{(i)}\right| + \bar{h}^{(i)}(y^{(i)}, y_{\mathcal{K}_i})\Big)d\tau,\end{aligned} \tag{21}$$

where $\bar{x}^{(i)}$ is a known bound such that $|x^{(i)}(t)| \leq \bar{x}^{(i)}$ for all t (see Assumption 2), and $\rho^{(i)} > 0$, $\lambda^{(i)} > 0$ are respectively

selected such that $e^{A_L^{(i)}t} \leq \rho^{(i)} e^{-\lambda^{(i)}t}$, for all t. The ARR $\mathcal{E}^{(i)}$ is robust to system disturbances and noise, implying that $\mathcal{M}^{(i)}$ does not raise false alarms.

3.2 Local Fault Identification

The primary goal of this step is the identification of the type of the fault that may have affected the local system, i.e. actuator or sensor fault or both. This is realized using two identification modules, $\mathcal{I}_a^{(i)}$ and $\mathcal{I}_s^{(i)}$ and an aggregation module $\mathcal{A}^{(i)}$ for fault isolation. The fault identification decision logic of $\mathcal{I}_a^{(i)}$ and $\mathcal{I}_s^{(i)}$ is based on two ARRs, $\mathcal{E}_a^{(i)}$ and $\mathcal{E}_s^{(i)}$ described by

$$\mathcal{E}_a^{(i)}: \ |\varepsilon_{y_a}^{(i)}(t)| \leq \bar{\varepsilon}_{y_a}^{(i)}(t), \quad \mathcal{E}_s^{(i)}: \ |\varepsilon_{y_s}^{(i)}(t)| \leq \bar{\varepsilon}_{y_s}^{(i)}(t), \tag{22}$$

where $\varepsilon_{y_a}^{(i)}(t)$, $\varepsilon_{y_s}^{(i)}(t)$ are the residuals generated by $\mathcal{I}_a^{(i)}$ and $\mathcal{I}_s^{(i)}$ respectively and $\bar{\varepsilon}_{y_a}^{(i)}(t)$, $\bar{\varepsilon}_{y_s}^{(i)}(t)$ are their corresponding adaptive thresholds. Due to the design of the adaptive thresholds, which is presented next: (i) the ARR $\mathcal{E}_a^{(i)}$ is guaranteed to be satisfied when $f_s^{(j)}$=0 for all $j \in \{\mathcal{K}_i \cup \{i\}\}$, and (ii) $\mathcal{E}_s^{(i)}$ is guaranteed to be satisfied when $f_s^{(j)}$=0 for all $j \in \mathcal{K}_i$ and $f_a^{(i)} = 0$. Therefore, as long as $\mathcal{E}_a^{(i)}$ is satisfied, $\mathcal{I}_a^{(i)}$ infers the occurrence of local actuator fault. As long as $\mathcal{E}_s^{(i)}$ is satisfied, $\mathcal{I}_s^{(i)}$ infers the occurrence of local sensor fault .

The outputs of the modules $\mathcal{I}_a^{(i)}$ and $\mathcal{I}_s^{(i)}$ are two local boolean decision functions $I_a^{(i)}$, $I_s^{(i)}$, defined as

$$I_a^{(i)}(t) = \begin{cases} 0, & t < t_{I_a}^{(i)} \\ 1, & t \geq t_{I_a}^{(i)} \end{cases}, \quad I_s^{(i)}(t) = \begin{cases} 0, & t < t_{I_s}^{(i)} \\ 1, & t \geq t_{I_s}^{(i)} \end{cases}, \tag{23}$$

where $t_{I_a}^{(i)} \triangleq inf\{t \geq t_D^{(i)} : |\varepsilon_{y_a}^{(i)}(t)| > \bar{\varepsilon}_{y_a}^{(i)}(t)\}$ and $t_{I_s}^{(i)} \triangleq inf\{t \geq t_D^{(i)} : |\varepsilon_{y_s}^{(i)}(t)| > \bar{\varepsilon}_{y_s}^{(i)}(t)\}$.

Residual Generation: The residuals associated with the modules $\mathcal{I}_a^{(i)}$ and $\mathcal{I}_s^{(i)}$ are defined as,

$$\varepsilon_{y_a}^{(i)}(t) = y^{(i)}(t) - \hat{x}_a^{(i)}(t), \tag{24}$$

$$\varepsilon_{y_s}^{(i)}(t) = y^{(i)}(t) - \hat{x}_s^{(i)}(t) - \hat{f}_s^{(i)}, \tag{25}$$

where $\hat{x}_a^{(i)}$ and $\hat{x}_s^{(i)}$ are both estimations of $x^{(i)}$, and $\hat{f}_s^{(i)}$ is the estimation of the sensor fault $f_s^{(i)}$. Assuming only local faults (i.e. $f_s^{(j)}$=0, $\forall\ j \in \mathcal{K}_i$), the state estimate $\hat{x}_a^{(i)}$ and $\hat{x}_s^{(i)}$ are computed based on the following distributed adaptive nonlinear estimation schemes

$$\begin{aligned} \dot{\hat{x}}_a^{(i)}(t) = & A^{(i)}\hat{x}_a^{(i)}(t) + g^{(i)}(y^s(t), y^{(i)}(t))(u_c^{(i)}(t) + \hat{f}_a^{(i)}(t)) \\ & + h^{(i)}(y^{(i)}(t), y_{\mathcal{K}_i}(t)) + \eta^{(i)}(d^{(i)}(t)) \\ & + L_a^{(i)}\varepsilon_{y_a}^{(i)}(t) + \Omega_a^{(i)}(t)\dot{\hat{f}}_a^{(i)}(t), \end{aligned} \tag{26}$$

$$\dot{\Omega}_a^{(i)}(t) = A_{L_a}^{(i)}\Omega_a^{(i)}(t) + g^{(i)}(y^s(t), y^{(i)}(t)), \tag{27}$$

$$\dot{\hat{f}}_a^{(i)}(t) = \gamma_a^{(i)}\Omega_a^{(i)}(t)\mathcal{D}^{(i)}\left[\varepsilon_{y_a}^{(i)}(t)\right], \tag{28}$$

$$\begin{aligned} \dot{\hat{x}}_s^{(i)}(t) = & A^{(i)}\hat{x}_s^{(i)}(t) + g^{(i)}(y^s(t), y^{(i)}(t) - \hat{f}_s^{(i)}(t))u_c^{(i)}(t) \\ & + h^{(i)}(y^{(i)}(t) - \hat{f}_s^{(i)}(t), y_{\mathcal{K}_i}(t)) + \eta^{(i)}(d^{(i)}(t)) \\ & + L_s^{(i)}\varepsilon_{y_s}^{(i)}(t) + \Omega_s^{(i)}(t)\dot{\hat{f}}^{(i)}(t), \end{aligned} \tag{29}$$

$$\dot{\Omega}_s^{(i)}(t) = A_{L_s}^{(i)}\Omega_s^{(i)}(t) - L_s^{(i)} + \sigma^{(i)}u_c^{(i)}(t), \tag{30}$$

$$\dot{\hat{f}}_s^{(i)}(t) = \gamma_s^{(i)}(\Omega_s^{(i)}(t) + 1)\mathcal{D}^{(i)}\left[\varepsilon_{y_s}^{(i)}(t)\right], \tag{31}$$

where $L_a^{(i)}$, $L_s^{(i)}$ are the estimation gains, such that $A_{L_a}^{(i)} \triangleq A^{(i)} - L_a^{(i)}$, $A_{L_s}^{(i)} \triangleq A^{(i)} - L_s^{(i)}$ are stable. The term $\hat{f}_a^{(i)}$ and $\hat{f}_s^{(i)}$ are the estimation of the fault $f_a^{(i)}$ and $f_s^{(i)}$, respectively. The positive constants $\gamma_a^{(i)}$, $\gamma_s^{(i)}$ are the learning rates of the adaptive laws in (28) and (31), and $\Omega_a^{(i)}$, $\Omega_s^{(i)}$ are filtering terms necessary for ensuring the stability of the adaptive schemes. Note that $\hat{x}_a^{(i)}(t_D^{(i)}) = 0$, $\hat{x}_s^{(i)}(t_D^{(i)}) = 0$, $\hat{f}_a^{(i)}(t_D^{(i)}) = 0$ and $\hat{f}_s^{(i)}(t_D^{(i)}) = 0$, $\Omega_a^{(i)}(t_D^{(i)}) = 0$ and $\Omega^{(i)}(t_D^{(i)}) = 0$, where $t_D^{(i)}$ is the detection time. The term $\mathcal{D}^{(i)}[.]$ represents the dead-zone operator

$$\mathcal{D}^{(i)}\left[\varepsilon_{y_\star}^{(i)}(t)\right] = \begin{cases} 0, & \text{if } D^{(i)}(t) = 0 \\ \varepsilon_{y_\star}^{(i)}(t), & \text{if } D^{(i)}(t) = 1 \end{cases}, \tag{32}$$

where $\varepsilon_{y_\star}^{(i)}$ represents either $\varepsilon_{y_a}^{(i)}$ in (28) or $\varepsilon_{y_s}^{(i)}$ in (31) and $D^{(i)}$ is defined in (11).

Computation of Adaptive Thresholds: Assuming $f_s^{(i)} = 0$ in (8), the residual in (24) can be expressed as,

$$\varepsilon_{y_a}^{(i)}(t) = \varepsilon_{x_a}^{(i)}(t) + n^{(i)}(t), \tag{33}$$

where $\varepsilon_{x_a}^{(i)}(t) = x^{(i)}(t) - \hat{x}_a^{(i)}(t)$ is the state estimation error. By using (5) and (26) and replacing $g^{(i)}\left(y^s, y^{(i)}\right)\tilde{f}_a^{(i)}$ with $(\dot{\Omega}_a^{(i)}(t) - A_L^{(i)}\Omega_a^{(i)}(t))\tilde{f}_a^{(i)}$ (see (27)), and after performing some mathematical manipulations, $\varepsilon_{x_a}^{(i)}$ satisfies

$$\varepsilon_{x_a}^{(i)} = \Omega_a^{(i)}\tilde{f}_a^{(i)} + \tilde{\varepsilon}_{x_a}^{(i)}, \tag{34}$$

$$\begin{aligned} \dot{\tilde{\varepsilon}}_{x_a}^{(i)} = & A_{L_a}^{(i)}\tilde{\varepsilon}_{x_a}^{(i)} + \tilde{g}^{(i)}(x^s, x^{(i)}, y^s, y^{(i)})u_c^{(i)} \\ & + \tilde{h}^{(i)}(x^{(i)}, x_{\mathcal{K}_i}, y^{(i)}, y_{\mathcal{K}_i}) - L_a^{(i)}n^{(i)} + r^{(i)}. \end{aligned} \tag{35}$$

where $\tilde{f}_a^{(i)}(t) = f_a^{(i)} - \hat{f}_a^{(i)}(t)$ is the actuator fault estimation error and $\dot{\tilde{f}}_a^{(i)}(t) = -\dot{\hat{f}}_a^{(i)}(t)$.

The residual in (25) can be expressed as,

$$\varepsilon_{y_s}^{(i)}(t) = \varepsilon_{x_s}^{(i)}(t) + \tilde{f}_s^{(i)}(t) + n^{(i)}(t), \tag{36}$$

where $\varepsilon_{x_s}^{(i)}(t) = x^{(i)}(t) - \hat{x}_s^{(i)}(t)$ is the state estimation error and $\tilde{f}_s^{(i)}(t) = f_s^{(i)} - \hat{f}_s^{(i)}(t)$ is the sensor fault estimation error with $\dot{\tilde{f}}_s^{(i)}(t) = -\dot{\hat{f}}_s^{(i)}(t)$. Assuming that $f_a^{(i)} = 0$ in (9), the dynamics of $\varepsilon_{x_s}^{(i)}$ can be described by

$$\begin{aligned} \dot{\tilde{\varepsilon}}_{x_s}^{(i)} = & A_{L_s}^{(i)}\varepsilon_{x_s}^{(i)} + \Omega_s^{(i)}\dot{\tilde{f}}_s^{(i)} + \sigma^{(i)}\tilde{f}_s^{(i)}u_c^{(i)} - L_s^{(i)}\tilde{f}_s^{(i)} \\ & + p^{(i)}\sum_{j\in\mathcal{K}_i} A_{d_{ij}}\left(\mu^{(i)}(x^{(i)}, x^{(j)}) - \mu^{(i)}(y^{(i)} - \hat{f}_s^{(i)}, y^{(j)})\right) \\ & - \sum_{j\in\mathcal{K}_i}\frac{a_{z_{ij}}}{C_{z_i}}n^{(j)} + \tilde{g}^{(i)}(x^s, x^{(i)}, y^s, y^{(i)})u_c^{(i)} - L_s^{(i)}n^{(i)} + r^{(i)}. \end{aligned} \tag{37}$$

Using (30), (37) can be re-written as:

$$\varepsilon_{x_s}^{(i)} = \Omega_s^{(i)}\tilde{f}_s^{(i)} + \tilde{\varepsilon}_{x_s}^{(i)}, \tag{38}$$

$$\begin{aligned} \dot{\tilde{\varepsilon}}_{x_s}^{(i)} = & A_{L_s}^{(i)}\tilde{\varepsilon}_{x_s}^{(i)} + p^{(i)}\sum_{j\in\mathcal{K}_i} A_{d_{ij}}\left(\mu^{(i)}(x^{(i)}, x^{(j)}) - \mu^{(i)}(y^{(i)} - \hat{f}_s^{(i)}, y^{(j)})\right) \\ & - \sum_{j\in\mathcal{K}_i}\frac{a_{z_{ij}}}{C_{z_i}}\overline{n}^{(j)} + \tilde{g}^{(i)}(x^s, x^{(i)}, y^s, y^{(i)})u_c^{(i)} - L_s^{(i)}n^{(i)} + r^{(i)}. \end{aligned} \tag{39}$$

Note that under the assumption of zero system disturbance and measurement noise, the errors $\tilde{\varepsilon}_{x_a}^{(i)}$ in (35) and $\tilde{\varepsilon}_{x_s}^{(i)}$ in (38) converge. If we also assume the persistence of excitation of the filters $\Omega_a^{(i)}$ and $\Omega_s^{(i)}$ in (27) and (30) respectively, then $\tilde{f}_a^{(i)}$ and $\tilde{f}_s^{(i)}$ converge as well.

Taking into account (22), the adaptive threshold $\bar{\varepsilon}_{y_a}^{(i)}(t)$ is computed by using (34) and the solution of (35) in (33), and the adaptive threshold and $\bar{\varepsilon}_{y_s}^{(i)}(t)$ is computed by using (38) and the solution of (39) in (36), and bounding each term, based on Assumptions 1 and 2 and the following assumption:

Assumption 3: The actuator and sensor faults $f_a^{(i)}$, $f_s^{(i)}$ are bounded; i.e. $|f_a^{(i)}(t)| \leq \overline{f}_a^{(i)}$ and $|f_s^{(i)}(t)| \leq \overline{f}_s^{(i)}$.

The adaptive thresholds $\bar{\varepsilon}_{y_a}^{(i)}(t)$ and $\bar{\varepsilon}_{y_a}^{(i)}(t)$ are defined as:

$$\begin{aligned}\bar{\varepsilon}_{y_a}^{(i)} &= \delta_a^{(i)} e^{\left(-\xi_a^{(i)}\left(t-T_D^{(i)}\right)\right)}\overline{x}^{(i)} + \left|\Omega_a^{(i)}(t)\right|\left(\left|\hat{f}_a^{(i)}\right| + \overline{f}_a^{(i)}\right) + \overline{n}^{(i)} \\ &+ \int_{T_D^{(i)}}^{t} \delta_a^{(i)} e^{\left(-\xi_a^{(i)}(t-\tau)\right)}\Big(\left|L_a^{(i)}\right|\overline{n}^{(i)} + \overline{r}^{(i)} + \overline{g}^{(i)}(\overline{n}^{(i)}, \overline{n}^{s})\left|u_c^{(i)}\right| \\ &+ \overline{g}^{(i)}(\overline{n}^{(i)}, \overline{n}^{s})\overline{f}_a^{(i)} + \overline{h}^{(i)}(y^{(i)}, y_{\mathcal{K}_i})\Big)d\tau, \end{aligned} \tag{40}$$

$$\begin{aligned}\bar{\varepsilon}_{y_s}^{(i)}(t) &= \delta_s^{(i)} e^{(-\xi_s^{(i)}(t-T_D^{(i)}))}\overline{x}^{(i)} + \left(\left|\Omega_s^{(i)}(t)\right| + 1\right)\left(\left|\hat{f}_s^{(i)}\right| + \overline{f}_s^{(i)}\right) + \overline{n}^{(i)} \\ &+ \int_{T_D^{(i)}}^{t} \delta_s^{(i)} e^{(-\xi_s^{(i)}(t-\tau))}\Bigg(\overline{g}^{(i)}(\overline{n}^{s}, \overline{n}^{(i)})\left|u_c^{(i)}(\tau)\right| + \sum_{j\in\mathcal{K}_i}\frac{a_{z_{ij}}}{C_{z_i}}\overline{n}^{(j)} \\ &+ \left|L_s^{(i)}\right|\overline{n}^{(i)} + \overline{r}^{(i)} + p^{(i)}\sum_{j\in\mathcal{K}_i} A_{d_{ij}}\overline{\mu}_f^{(i)}(y^{(i)} - \hat{f}_s^{(i)}, y^{(j)})\Bigg)d\tau, \end{aligned} \tag{41}$$

where $\overline{g}^{(i)}$ and $\overline{h}^{(i)}$ are defined in (18) and (19) respectively and $\delta_a^{(i)}$, $\xi_a^{(i)}$ and $\delta_s^{(i)}$, $\xi_s^{(i)}$ are selected such that $e^{(A_{L_a}^{(i)}t)} \leq \delta_a^{(i)} e^{(-\xi_a^{(i)}t)}$ and $e^{(A_{L_s}^{(i)}t)} \leq \delta_s^{(i)} e^{(-\xi_s^{(i)}t)}$, respectively, and $\overline{\mu}_f^{(i)}(y^{(i)} - \hat{f}_s^{(i)}, y^{(j)})$ is defined in Appendix A.

It is noted that based on the design of $\bar{\varepsilon}_{y_a}^{(i)}$ and $\bar{\varepsilon}_{y_s}^{(i)}$, the ARRs $\mathcal{E}_a^{(i)}$ and $\mathcal{E}_s^{(i)}$ defined in (22) are respectively insensitive to $f_a^{(i)}$ and $f_s^{(i)}$.

Remark 3.1. The distributed fault detection process is applied before the local fault identification in order to reduce the computational effort of the agent $\mathcal{M}^{(i)}$ during the healthy operation of the system that may be long. Particularly, as shown in Section 3.1, one non-adaptive estimator is used, generating a single residual that is compared to its corresponding threshold. After the first time of fault detection, the local identification process is continuously active.

3.3 Distributed Fault isolation

The decisions of the two modules $\mathcal{I}_a^{(i)}$ and $\mathcal{I}_s^{(i)}$ are collected by the aggregation module $\mathcal{A}^{(i)}$ that processes them in combination. The decisions $I_a^{(i)}$, $I_s^{(i)}$ constitute the observed fault pattern defined as

$$I^{(i)}(t) = [I_a^{(i)}(t), I_s^{(i)}(t)]^{\top}. \tag{42}$$

This pattern is compared to the columns of the fault signature matrix denoted by $F^{(i)}$ shown in Table 1, where the term $f_{\mathcal{K}_i}$ collectively amounts for the sensor faults propagated by the neighboring agents due to the exchange of information, and $\mathcal{F}_{\mathcal{K}_i}$ represents all the combinations of local and propagated faults. The element of $F^{(i)}$ equals to 0 when the corresponding ARR has been designed to be insensitive to this fault, and equals to 1 otherwise.

The outcome of the comparison between the observed pattern $I^{(i)}(t)$ to the columns of $F^{(i)}$ is the diagnosis set $\Delta_I^{(i)}(t)$, including the diagnosed fault combinations that may have occurred. When $I^{(i)}(t) = [1,1]^{\top}$, the diagnosis set contains more than one combinations, and the distributed fault isolation process is activated in order to decide if only local faults have occurred or also sensor faults have been propagated. Otherwise, it is inferred that a single local fault has occurred which is either actuator fault (if $I^{(i)}(t) = [0,1]^{\top}$), or sensor fault (if $I^{(i)}(t) = [1,0]^{\top}$). The decision about the propagation of sensor faults is defined as

$$I_{\mathcal{K}_i}^{(i)}(t) = \begin{cases} 0, \text{ if } \left(f_s^{(i)} \notin \Delta_I^{(i)}(t) \;\&\; f_{\mathcal{K}_i} \notin \Delta_I^{(i)}(t)\right) \text{ or } D^{(i)}(t) = 0 \\ 1, \qquad \text{if } f_s^{(i)} \in \Delta_I^{(i)}(t) \text{ or } f_{\mathcal{K}_i} \in \Delta_I^{(i)}(t) \end{cases}.$$

Table 1. Fault isolation signature matrix $F^{(i)}$

	$f_a^{(i)}$	$f_s^{(i)}$	$\{f_a^{(i)}, f_s^{(i)}\}$	$f_{\mathcal{K}_i}$	$\mathcal{F}_{\mathcal{K}_i}$
$\mathcal{E}_a^{(i)}$	0	1	1	1	1
$\mathcal{E}_s^{(i)}$	1	0	1	1	1

When $I_{\mathcal{K}_i}^{(i)} = 1$, the agent $\mathcal{M}^{(i)}$ requests from all neighboring agents $\mathcal{M}^{(j)}$, $j \in \mathcal{K}_i$, to transmit their decisions $I_{\mathcal{K}_j}^{(j)}(t)$, creating the observed pattern of propagated sensor faults, determined as

$$I_{\mathcal{K}_i}(t) = \left[I_{\mathcal{K}_j}^{(j)}(t) : j \in \mathcal{K}_i \cup \{i\}\right]^{\top}. \tag{43}$$

This pattern is compared to the columns of a fault signature matrix denoted by $F_{\mathcal{K}_i}$ with $c = \text{card}(\mathcal{K}_i) + 1$ rows and $2^c - 1$ columns. Each row corresponds to the ARR $\mathcal{E}_{\mathcal{K}_i} = \mathcal{E}_a^{(i)} \cup \mathcal{E}_s^{(i)}$ and each column corresponds to a combination of sensors faults in the set $f_s^{(i)} \cup f_{\mathcal{K}_i}$. The element (p,q) of $F_{\mathcal{K}_i}$, $p \in \{1,\ldots,c\}$, $q \in \{1,\ldots,2^c-1\}$ equals to 0 when the q-th ARR is structurally insensitive to the fault combination q. If the q-th fault combination includes the sensor fault $f_s^{(i)}$ and the p-th row corresponds to the ARR $\mathcal{E}_{\mathcal{K}_i}$, then the element (p,q) equals to 1, since $f_s^{(i)}$ is a local sensor fault for $\mathcal{E}_{\mathcal{K}_i}$. If the q-th fault combination includes only faults $f_s^{(j)}$, $j \in \mathcal{K}_i$ and the p-th row corresponds to the ARR $\mathcal{E}_{\mathcal{K}_i}$, then the element (p,q) is set to the symbol $*$, which represents either 1 or 0 [Reppa et al. (2016)]. An example of the matrix $F_{\mathcal{K}_i}$ is shown in Table 2 of the simulation example. The outcome of the comparison is the diagnosis set $\Delta_{\mathcal{K}_i}$ which includes the possible fault combinations of propagated sensor faults. If $f_s^{(j)} \notin \Delta_{\mathcal{K}_i}$ for all $j \in \mathcal{K}_i$, then the agent $\mathcal{M}^{(i)}$ infers the occurrence of local faults, while if there is at least one fault $f_s^{(q)} \notin \Delta_{\mathcal{K}_i}$, $q \in \mathcal{K}_i$ then the agent $\mathcal{M}^{(i)}$ infers that sensor faults may have been propagated from the agents $\mathcal{M}^{(j)}$ $j \in \{\mathcal{K}_i \setminus \{q\}\}$.

4. SIMULATION RESULTS

In this section, we illustrate the application of the distributed diagnostic scheme, presented in Section 3 to a 5-zone HVAC system whose down-view is presented with solid black lines in Fig. 1. Based on Fig. 1, we define the following index sets $\mathcal{K}_1 = \{2,3\}$, $\mathcal{K}_2 = \{1,3,5\}$, $\mathcal{K}_3 = \{1,2,4,5\}$, $\mathcal{K}_4 = \{3,5\}$, $\mathcal{K}_5 = \{2,3,4\}$. The parameters of each subsystem are: a_{z_i}=740 KJ/h°C, $a_{z_{ij}}$=50 KJ/h°C, a_{st}=12 KJ/kg°C, a_{sz}=0.6 KJ/h°C, C_{st}=8370 kJ/°C, C_p=1.004 kJ/kgK, C_v=0.717 kJ/kgK, ρ_{air}=1.22 kg/m^3, C_{z_1}=30, C_{z_2}=58, C_{z_3}=55, C_{z_4}= C_{z_5}=27 kJ/°C, $U_{i,max}$=3700 kg/h, p=2.5, $U_{st,max}$=27.36×10^5

kg/h, ΔT_{max}=45 °C, A_{w_i}=120 m^2, h=8.29 W/m^2°C, $A_{d_{ij}}$=1.95m^2, d_1^s=$d_1^{(i)}$=$d_2^{(i)}$=10°C, d_2^s=5°C. The modeling uncertainties are modelled as r^s= 10%$d_1^s \sin(0.1t)$ (°C/h), $r^{(i)}$= 10%$d_1^{(i)} \sin(0.1t)$ (°C/h), $i \in \{1,\dots,5\}$. The desired temperatures are selected as y_{ref}^s = 55°C, $y_{ref}^{(i)}$ = 24°C, $\forall i$. The design parameters of the monitoring agents are: $\overline{n}^s$ = 3%y_{ref}^s, $\overline{n}^{(i)}$ = 3%$y_{ref}^{(i)}$, $\overline{r}^s$ = 10%d_1^s, $\overline{r}^{(i)}$ = 10%$d_1^{(i)}$, $L^{(i)}$ = 15, $L_a^{(i)}$=4, $L_s^{(i)}$=22, $\rho^{(i)}$ = 1.1, $\lambda^{(i)}$ = 25, $\delta_a^{(i)}$=1.1, $\xi_a^{(i)}$=35, $\delta_s^{(i)}$=1.1, $\xi_s^{(i)}$=15, $\gamma_a^{(i)}$=5, $\gamma_s^{(i)}$=32, $\overline{f}_a^{(i)}$=1.5, $\overline{f}_s^{(i)}$=6, $\overline{x}^{(i)}$=20°C, $\forall i$.

To illustrate the decision-making process of the agents a multiple fault scenario is performed. Specifically, two consecutive faults occur in zone 1 and they have been simulated such that $f_s^{(1)}$ = 20%$y_{ref}^{(1)}$ at t = 0.4h and $f_a^{(1)}$ = −25%$u_n^{(1)}$ at t = 0.6h, with $u_n^{(1)}$ = 0.2 where $u_n^{(1)}$=$u^{(1)}$ in steady state and healthy conditions.

Fig. 3 presents the fault detection process of the agents $\mathcal{M}^{(i)}$, $i \in \{1,\dots,5\}$. The agent $\mathcal{M}^{(1)}$ detects a fault at the time instant $t_D^{(1)}$ = 0.4h. Note that, the activation of the local identification process is realized at the first time instant that the agents detect faults. On the contrary, none of the remainder agents ($\mathcal{M}^{(i)}$, $i \in \{2,\dots,5\}$) detected any fault.

Fig. 4 presents the simulation results of the local identification process of $\mathcal{M}^{(1)}$. At the time instant 0.4871h the aggregation module $\mathcal{A}^{(1)}$ collects the decisions of the two identification modules and compares the observed pattern $I^{(1)}(0.4871) = [I_a^{(1)}(0.4871), I_s^{(1)}(0.4871)]^\top = [1,0]^\top$ to the columns of Table 1. The agent $\mathcal{M}^{(1)}$ decides that a single sensor fault has occurred in zone 1. The local fault identification continues being active, as well as the comparison of the observed pattern $I^{(1)}(t)$ to the columns of Table 1. As shown in Fig. 4 at the time instant 0.6551h, the pattern $I^{(1)}$ becomes $I^{(1)}(0.6551) = [I_a^{(1)}(0.6551), I_s^{(1)}(0.6551)]^\top = [1,1]^\top$, leading to $\Delta_I^{(1)}(0.6551) = \left\{\left\{f_a{}^{(1)}, f_s{}^{(1)}\right\}, f_{\mathcal{K}_1}{}^{(1)}, \mathcal{F}_{\mathcal{K}_1}{}^{(1)}\right\}$ and $I_{\mathcal{K}_1}^{(1)}(0.6551) = 1$. Then, the agent $\mathcal{M}^{(1)}$ requests the transmission of the decisions $I_{\mathcal{K}_2}^{(2)}$ and $I_{\mathcal{K}_3}^{(3)}$ of the agents $\mathcal{M}^{(2)}$ and $\mathcal{M}^{(3)}$ respectively, creating the observed pattern $I_{\mathcal{K}_1}(0.6551) = [I_{\mathcal{K}_1}^{(1)}(0.6551), I_{\mathcal{K}_2}^{(2)}(0.6551), I_{\mathcal{K}_3}^{(3)}(0.6551)]^\top = [1,0,0]^\top$. This pattern is compared to the columns of the matrix shown in Table 2, leading to the diagnosis set $\Delta_{\mathcal{K}_1} = \{f_s^{(1)}\}$. Based on this set, the agent $\mathcal{M}^{(1)}$ excludes the propagation of sensor faults and infers the occurrence of local actuator and sensor faults.

Table 2. Distributed fault signature matrix $F_{\mathcal{K}_1}$ ($f_s^{(1,2)} = \{f_s^{(1)}, f_s^{(2)}\}$, $f_s^{(1,3)} = \{f_s^{(1)}, f_s^{(3)}\}$, $f_s^{(2,3)} = \{f_s^{(2)}, f_s^{(3)}\}$, $f_s^{(1,2,3)} = \{f_s^{(1)}, f_s^{(2)}, f_s^{(3)}\}$)

	$f_s^{(1)}$	$f_s^{(2)}$	$f_s^{(3)}$	$f_s^{(1,2)}$	$f_s^{(1,3)}$	$f_s^{(2,3)}$	$f_s^{(1,2,3)}$
$\mathcal{E}_{\mathcal{K}_1}$	1	*	*	1	1	*	1
$\mathcal{E}_{\mathcal{K}_2}$	*	1	*	1	*	1	1
$\mathcal{E}_{\mathcal{K}_3}$	*	*	1	*	1	1	1

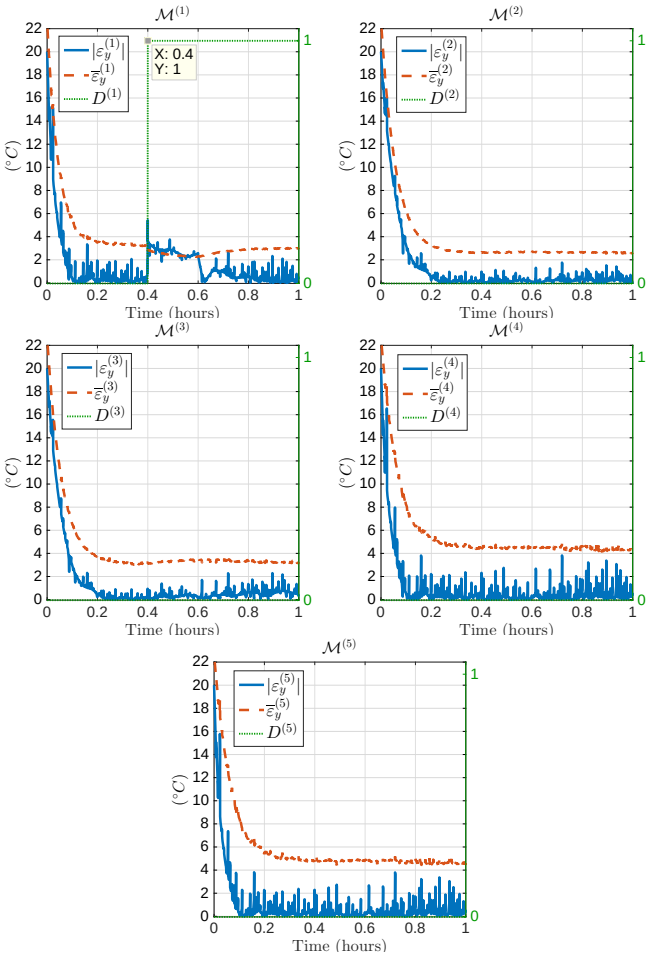

Fig. 3. Fault detection process of agents $\mathcal{M}^{(i)}$, $i \in \{1,2,3,4,5\}$.

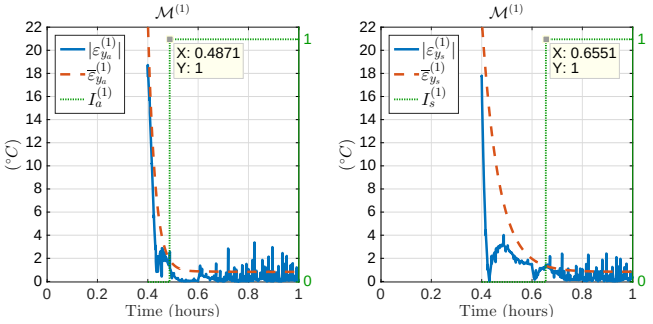

Fig. 4. Local fault identification of $\mathcal{I}_a^{(1)}$ (left) and $\mathcal{I}_s^{(1)}$ (right).

5. CONCLUSION

This paper presented a distributed fault diagnosis (FD) methodology for isolating actuator and sensor faults in a multi-zone HVAC system. The proposed architecture relies on the deployment of several distributed monitoring agents, which are allowed to exchange information. Every agent is designed to detect the presence of faults, identify the type and infer the number and location (local or propagated faults).

Appendix A

The bound $\overline{\mu}^{(i)}$ can be determined by using inclusion functions and applying interval analysis in order to compute a time-varying interval for the function $\mu^{(i)}(x^{(i)}, x^{(j)})$ defined in (20); i.e., belongs to the following interval

$$\mu^{(i)}(x^{(i)}(t), x^{(j)}(t)) \in [\underline{\chi}^{(i)}(t), \overline{\chi}^{(i)}(t)], \ \forall t, \quad \text{(A.1)}$$

where $\underline{\chi}^{(i)}(t)$, $\overline{\chi}^{(i)}(t)$ are the lower and upper time-varying endpoints. Let us define $\tilde{\mu}^{(i)}(x^{(i)}, x^{(j)}, y^{(i)}, y^{(j)}) = \mu^{(i)}(x^{(i)}, x^{(j)}) - \mu^{(i)}(y^{(i)}, y^{(j)})$. Using (A.1) and applying interval arithmetic results in

$$\tilde{\mu}^{(i)}(x^{(i)}, x^{(j)}, y^{(i)}, y^{(j)}) \in \left[\underline{\chi}^{(i)} - \upsilon^{(i)}, \overline{\chi}^{(i)} - \upsilon^{(i)})\right] \quad \text{(A.2)}$$

with $\upsilon^{(i)}(t) = \mu^{(i)}(y^{(i)}(t), y^{(j)}(t))$. The upper bound that satisfies $|\tilde{\mu}^{(i)}(x^{(i)}, x^{(j)}, y^{(i)}, y^{(j)})| \leq \overline{\mu}^{(i)}(y^{(i)}, y^{(j)})$ is computed as

$$\overline{\mu}^{(i)}(y^{(i)}, y^{(j)}) = \max(|\underline{\chi}^{(i)} - \upsilon^{(i)}|, |\overline{\chi}^{(i)} - \upsilon^{(i)}|). \quad \text{(A.3)}$$

Given (20), we define $\chi_1 = \sqrt{|x^{(j)} - x^{(i)}|}$ and $\chi_2 = \text{sgn}(x^{(j)} - x^{(i)})\max(x^{(i)}, x^{(j)})$. Based on (8) and Assumption 1, under healthy conditions (i.e. $f_s^{(i)} = 0$ for all i), we have $x^{(i)}(t) \in \left[y^{(i)}(t) - \overline{n}^{(i)}, y^{(i)}(t) + \overline{n}^{(i)}\right]$, $\forall i$ and $\forall t$. Taking into account the monotonicity of χ_1 and applying interval arithmetic, we obtain

$$\chi_1 \in \left[\underline{\chi}_1^{(i)}, \overline{\chi}_1^{(i)}\right] \quad \text{(A.4)}$$

$$\left[\underline{\chi}_1^{(i)}, \overline{\chi}_1^{(i)}\right] = \begin{cases} \left[\sqrt{|\alpha(t) - \beta|}, \sqrt{|\alpha(t) + \beta|}\right], & \text{if } \alpha(t) > \beta \\ \left[\sqrt{|\alpha(t) + \beta|}, \sqrt{|\alpha(t) - \beta|}\right], & \text{if } \alpha(t) < -\beta \\ \left[0, \sqrt{\max\left(|\alpha(t) - \beta|, |\alpha(t) + \beta|\right)}\right], & \text{if } |\alpha(t)| \leq \beta \end{cases} \quad \text{(A.5)}$$

where $\alpha(t) = y^{(j)}(t) - y^{(i)}(t)$, $\beta = \overline{n}^{(j)} + \overline{n}^{(i)}$. Following similar steps, we have

$$\chi_2 \in \left[\underline{\chi}_2^{(i)}, \overline{\chi}_2^{(i)}\right] \quad \text{(A.6)}$$

$$\left[\underline{\chi}_2^{(i)}, \overline{\chi}_2^{(i)}\right] = \begin{cases} \left[y^{(j)}(t) - \overline{n}^{(j)}, y^{(j)}(t) + \overline{n}^{(j)}\right], & \text{if } \alpha(t) > \beta \\ \left[-y^{(i)}(t) - \overline{n}^{(i)}, -y^{(i)}(t) + \overline{n}^{(i)}\right], & \text{if } \alpha(t) < -\beta \\ [\min(W), \max(W)], & \text{if } |\alpha(t)| \leq \beta \end{cases} \quad \text{(A.7)}$$

$$W = \{-\min(w_1, w_2), -\max(w_1, w_2), \min(w_1, w_2), \max(w_1, w_2)\} \quad \text{(A.8)}$$

where $w_1 = y^{(j)}(t) + \overline{n}^{(j)}$ and $w_2 = y^{(i)}(t) + \overline{n}^{(i)}$. Taking into account that $\mu^{(i)}(x^{(i)}(t), x^{(j)}(t)) = \chi_2(t)\chi_1(t)$ and (A.4)-(A.6), it yields

$$\underline{\chi}^{(i)} = \min(\underline{\chi}_1^{(i)}\underline{\chi}_2^{(i)}, \underline{\chi}_1^{(i)}\overline{\chi}_2^{(i)}, \overline{\chi}_1^{(i)}\underline{\chi}_2^{(i)}, \overline{\chi}_1^{(i)}\overline{\chi}_2^{(i)}), \quad \text{(A.9)}$$

$$\overline{\chi}^{(i)} = \max(\underline{\chi}_1^{(i)}\underline{\chi}_2^{(i)}, \underline{\chi}_1^{(i)}\overline{\chi}_2^{(i)}, \overline{\chi}_1^{(i)}\underline{\chi}_2^{(i)}, \overline{\chi}_1^{(i)}\overline{\chi}_2^{(i)}). \quad \text{(A.10)}$$

Assuming the occurrence of the sensor fault $f_s^{(i)}$, while $f_s^{(j)} = 0$, $\forall\ j \in \mathcal{K}_i$, the bound $\overline{\mu}_f^{(i)}(y^{(i)} - \hat{f}_s^{(i)}, y^{(j)})$ can be computed based on (A.3)-(A.8). Specifically, is equal to the right-hand side of (A.3) where $\upsilon^{(i)} = \mu^{(i)}(y^{(i)} - \hat{f}_s^{(i)}, y^{(j)})$, and $\underline{\chi}^{(i)}$ and $\overline{\chi}^{(i)}$ are defined through (A.4)-(A.10) with β in (A.5) being equal to $\beta^{(i)} = \overline{n}^{(i)} + \overline{n}^{(j)} + \overline{f}_s^{(i)}$ and $\overline{n}^{(i)}$ in (A.7) and (A.8) being replaced with $\overline{n}^{(i)} + \overline{f}_s^{(i)}$.

REFERENCES

Blanke, M., Kinnaert, M., Lunze, J., and Staroswiecki, M. (2016). *Distributed Fault Diagnosis and Fault-Tolerant Control*. Springer.

Boem, F., Ferrari, R.M., Keliris, C., Parisini, T., and Polycarpou, M.M. (2017). A distributed networked approach for fault detection of large-scale systems. *IEEE Transactions on Automatic Control*, 62(1), 18–33.

Bonvini, M., Sohn, M.D., Granderson, J., Wetter, M., and Piette, M.A. (2014). Robust on-line fault detection diagnosis for HVAC components based on nonlinear state estimation techniques. *Applied Energy*, 124, 156 – 166.

Carbot-Rojas, D., Escobar, R., Gómez-Aguilar, J., López-López, G., and Olivares-Peregrino, V. (2015). Experimental validation of an actuator fault tolerant control system using virtual sensor: Application in a double pipe heat exchanger. *Chemical Engineering Research and Design*, 104, 400–408.

Darure, T., Yamé, J.J., and Hamelin, F. (2016). Fault-adaptive control of VAV damper stuck in a multizone building. In *Proceedings of 3rd Conference on Control and Fault-Tolerant Systems*, 164–170.

Keliris, C., Polycarpou, M.M., and Parisini, T. (2015). Distributed fault diagnosis for process and sensor faults in a class of interconnected input–output nonlinear discrete-time systems. *International Journal of Control*, 88, 1472–1489.

Lee, W.Y., House, J.M., and Kyong, N.H. (2004). Subsystem level fault diagnosis of a building's air-handling unit using general regression neural networks. *Applied Energy*, 77(2), 153–170.

Liang, J. and Du, R. (2007). Model-based Fault Detection and Diagnosis of HVAC systems using Support Vector Machine method. *International Journal of Refrigeration*, 30(6), 1104–1114.

Papadopoulos, P.M., Reppa, V., Polycarpou, M.M., and Panayiotou, C.G. (2015a). Distributed Adaptive Estimation Scheme for Isolation of Sensor Faults in Multizone HVAC Systems. In *Proceedings of 9th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes*, 1146–1151.

Papadopoulos, P.M., Reppa, V., Polycarpou, M.M., and Panayiotou, C.G. (2015b). Distributed adaptive sensor fault tolerant control for smart buildings. In *Proceedings of 54th IEEE Conference on Decision and Control*, 3143–3148.

Reppa, V., Papadopoulos, P., Polycarpou, M.M., and Panayiotou, C.G. (2015). A distributed architecture for HVAC sensor fault detection and isolation. *IEEE Transactions on Control Systems Technology*, 23(4), 1323–1337.

Reppa, V., Polycarpou, M.M., and Panayiotou, C.G. (2016). Sensor Fault Diagnosis. *Foundations and Trends in Systems and Control*, 3(1-2), 1–248.

Wang, S., Zhou, Q., and Xiao, F. (2010). A system-level fault detection and diagnosis strategy for HVAC systems involving sensor faults. *Energy and Buildings*, 42(4), 477–490.

Weimer, J., Ahmadi, S.A., Araujo, J., Mele, F.M., Papale, D., Shames, I., Sandberg, H., and Johansson, K.H. (2012). Active actuator fault detection and diagnostics in HVAC systems. In *Proceedings of the Fourth ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, 107–114.

Witrant, E., Mocanu, S., and Sename, O. (2009). A hybrid model and MIMO control for intelligent buildings temperature regulation over WSN. *IFAC Proceedings Volumes*, 42(14), 420–425.